DOCUMENT RESUME

ED 355 252 TM 019 593

AUTHOR	Rudner, Lawrence M.; Shafer, Mary Morello
TITLE	Resampling: A Marriage of Computers and Statistics. ERIC/TM Digest.
INSTITUTION	ERIC Clearinghouse on Tests, Measurement, and Evaluation, Washington, DC.
SPONS AGENCY	Office of Educational Research and Improvement (ED), Washington, DC.
REPORT NO	EDO-TM-92-8
PUB DATE	Sep 92
CONTRACT	RI88062003
NOTE	4p.
AVAILABLE FROM	American Institutes for Research, 3333 K Street, N.W., Suite 300, Washington, DC 20007 (free).
PUB TYPE	Information Analyses - ERIC Clearinghouse Products (071)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	Computer Oriented Programs; Computer Uses in Education; *Educational Research; Elementary Secondary Education; *Estimation (Mathematics); Nonparametric Statistics; *Probability; *Research Methodology; Sampling; Statistical Distributions; *Statistics; Technological Advancement
IDENTIFIERS	Bootstrap Methods; Cross Validation; ERIC Digests; Jackknifing Technique; *Resampling Techniques; Research Replication

ABSTRACT

Advances in computer technology are making it possible for educational researchers to use simpler statistical methods to address a wide range of questions with smaller data sets and fewer, and less restrictive, assumptions. This digest introduces computationally intensive statistics, collectively called resampling techniques. Resampling is a process for estimating probabilities by conducting vast numbers of numerical experiments with the aid of high speed computers. Probability theory has freed researchers from the drudgery of repeated experiments, and resampling has now made it possible to analyze virtually any statistic in an easy to understand manner, without making assumptions about the distribution of the data. The following four techniques are important in resampling: (1) the bootstrap; (2) the jackknife; (3) cross-validation; and (4) balanced repeated replication. An example compares the grades of voucher and non-voucher students in an elementary school using the bootstrap. It is argued that resampling frees researchers from the assumption that the data conform to a bell-shaped curve and the need to focus on statistical measures with theoretical properties that can be analyzed mathematically. Critics question the resampling method itself and question the accuracy of the estimates that resampling yields. In some situations, resampling may be less accurate than conventional parametric methods. (SLD)

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

ED355252

RESAMPLING: A MARRIAGE OF COMPUTERS AND STATISTICS

Lawrence M. Rudner and Mary Morello Shafer, ERIC/TM

TM019593

BEST COPY AVAILABLE

ERIC Clearinghouse on Tests, Measurement, and Evaluation September 1992 EDO-TM-92-8

Resampling: a marriage of computers and statistics

Lawrence M. Rudner and Mary Morello Shafer, ERIC/TM

Suppose your superintendent asked you to determine whether voucher students are doing better than non-voucher students in your district's elementary schools. You might perform a simple *t* test or an analysis of variance to find your answer. You would report mean differences and probability levels. And if your superintendent is like most non-statisticians, he or she would accept the magic of statistics without questionning the validity of the assumptions made to use the *t* test.

Thanks to advances in computer technology, educational researchers are beginning to use simpler statistical methods. These techniques let us empiricially address a wider range of questions with smaller data sets and with fewer, less restrictive assumptions. Using such techniques, we can focus on reasoning and on understanding the data, not on complicated formulas and tables. The techniques promise to make statistics a useful, easily learned tool for educational policy makers and researchers.

This digest introduces computationally intensive statistics, collectively called *resampling techniques*. After defining these statistics, we'll use one technique to answer our opening question. We'll then present the arguments for and against resampling.

Resampling defined

Resampling is simply a process for estimating probabilities by conducting vast numbers of numerical experiments. Today, resampling is done with the aide of high speed computers.

In *Science News*, Peterson (1991) compares resampling techniques to the trial-and-error way gamblers once used to figure odds in card or dice games. Before the invention of probability theory, gamblers would deal out many hands of a card game to count the number of times a particular hand occurred. Thus, by experimentation, gamblers could figure the odds of getting a certain hand in their game.

Probability theory freed researchers from the drudgery of repeated experiments. With a few assumptions, researchers could address a wide range of topics. While the advances in statistics paved the way for elegant analysis, the costs came high:

- We could analyze only certain types of statistics, such as the mean and standard deviation.
- We had to make certain assumptions, like the normality assumption, about the underlying distribution.
- And researchers needed specialized training to apply, understand, and appreciate statistics.

But resampling techniques overcome all these limitations today:

- We can analyze virtually any statistic.
- We don't have to make any assumptions about the distribution of the data.
- And the techniques are easily to understand.

All resampling techniques rely on the computer to generate data sets from the original data. The techniques differ, however, in how they generate the data sets. Four techniques are important:

- the *bootstrap*, invented by Bradley Efron;
- the *jacknife*, invented by Maurice Quenouille and later developed by John W. Tukey;
- *cross-validation*, developed by Seymour Geisser, Mervyn Stone, and Grace G. Wahba; and
- *balanced repeated replication*, developed by Philip J. McCarthy.

A resampling example for education

Back to the question comparing the grades of voucher and non-voucher students: Using the bootstrap technique, we can empirically construct the distribution of mean grade differences for students in these two groups. If the observed difference is unusual, then we would reject the null hypothesis that grades are unrelated to voucher status.

For simplicity, let's assume that the district has 13 voucher students and 39 non-voucher students, and the mean difference is 10 standard score units. To empirically construct the distribution, we'd follow these steps:

1. Create a data base with all the student grades.

TM019593

2. Randomly sort the data base.
3. Compute the mean for the first 13 students.
4. Compute the mean for the other 39 students.
5. Record the test statistic—the absolute value of the mean difference.
6. Then repeat steps 2 though 5 many times.

That way, we'd get the distribution of mean differences when we randomly select students. The probability of observing a mean difference of 10 when everything is random is the proportion of experimental test statistics in step 5 that are greater than 10.

Noreen (1989) noted several striking aspects of this approach:

- Researchers make no assumptions about the distribution of grades (for example, no normality assumption).
- The data are not a random sample from some population.

Arguments *for* resampling

Diaconis and Efron (1983) argue that the resampling method free researchers from two limitations of conventional statistics: "the assumption that the data conform to a bell-shaped curve and the need to focus on statistical measures whose theoretical properties can be analyzed mathematically." Instead, Peterson says, this method "addresses a key problem in statistics: how to infer the 'truth' from a sample of data that may be incomplete or drawn from an ill-defined population."

The resampling method forces researchers to clarify the problem: With no formulas to fall back on, you have to explicitly define the question you want to answer. According to Simon and Bruce (1991), the method prevents researchers from "simply grabbing the formula for some test without understanding why they chose that test." As Peterson explains, instead of asking which formula to use, you "begin tackling such questions as what makes certain results statistically 'significant.' "

In *Scientific American,* Diaconis and Efron apply the bootstrap method to various types of problems and then compare the results from the bootstrap with the results from conventional statistical tests, including the correlation coefficient and principal components. Most of the time, the bootstrap method yielded the same answers that the more conventional methods did. Of course, the bootstrap may not give a true picture of every sample, just as conventional tests sometimes find deceptive answers to problems.

Because resampling techniques like the bootstrap are so easy to use and understand, Simon and Bruce advocate teaching these techniques to students first—that way, students learn how to translate their *"scientific* question into a *statistical* question." By learning how to think clearly about their problem, students won't "select their methods blindly."

They cite a study where one group of students learned resampling techniques, and the other learned conventional methods. The students taught the resampling techniques did much better solving statistical problems than the other students taught conventional methods. Further, the students who learned the resampling techniques enjoyed statistics, and their attitudes toward math improved during the course. However, the attitudes of the students who learned conventional techniques got worse during the course.

Arguments *against* resampling

Critics question the resampling method itself. They argue, as Stephen E. Fienberg says, that "you're trying to get something for nothing. You use the same numbers over and over again until you get an answer that you can't get any other way. In order to do that, you have to assume something, and you may live to regret that hidden assumption later on" (Peterson, 1991, p. 57).

Other critics question the accuracy of the estimates that resampling yields—if, for example, the researcher doesn't make enough experimental trials. In some situations, resampling may be less accurate than conventional parametric methods.

Additional reading

The classic introduction to this field:

Diaconis, P., and B. Efron. (1983). Computer-intensive methods in statistics. *Scientific American,* May, 116–130.

Numerous examples, as well as Basic, Pascal, and Fortran source code for conducting several resampling experiments:

Noreen, Eric. (1989). *Computer-intensive methods for testing hypothesis*. New York: Wiley.

Discusses using resampling methods to teach statistical concepts:

Peterson, I. (July 27, 1991). Pick a sample. *Science News,* 140, 56–58.

Low-cost IBM PC software for learning and applying resampling:

Simon, J. L. (1990). *Resampling stats: Probability and statistics, a radically different way*. University of Maryland, College of Business and Management.

Arguments for using and teaching resampling:

Simon, J. L., and P. Bruce. (1991). Resampling: A tool for everyday statistical work. *Chance*, 4(1), 22–32.

This publication was prepared with funding from the Office of Educational Research and Improvement, U.S. Department of Education, under contract number R-88-062003. The opinions expressed in this report do not necessarily reflect the position or policies of OERI or the Department of Education. ***Permission is granted to copy and distribute this Digest.***